Mol Gen Genet (1987) 208:107–113

MGG

Characterization of a null-allele for the Gy_4 glycinin gene from soybean*

Bernard J. Scallon, Craig D. Dickinson, and Niels C. Nielsen
USDA/ARS and the Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

Summary. A null allele for the Gy_4 glycinin gene from the cultivar Raiden was sequenced and compared with a functional Gy_4 gene from another cultivar. The results showed that the null phenotype probably resulted from a point mutation that changed the translation initiation codon from ATG to ATA. Transcripts of the mutant gene were detected in total RNA from seed, but the mRNAs did not become associated with polysomes as did functional Gy_4 message in the control cultivar. This was probably due to premature dissociation of mutant gy_4 mRNA from ribosomes due to nonsense codons during translation of an incorrect reading frame.

Key words: Glycinin mutant – Storage protein – DNA sequence

Introduction

Investigations into the molecular basis of null alleles afford an opportunity to learn what types of genetic lesions can affect gene expression. Although a number of null alleles have been identified, only a few have been characterized at the molecular level. These include null alleles for maize alcohol dehydrogenase (Bennetzen et al. 1984; Taylor and Walbot 1985), soybean seed lectin (Goldberg et al. 1983), and the α'-subunit of β-conglycinin (Ladin et al. 1984). The mutations in these alleles are due to insertions or deletions that affected transcription.

Null alleles have also been reported among members of the gene family that encode glycinin subunits from soybeans. The best characterized of the glycinin null phenotypes is found in the cultivar Raiden. It lacks the $A_5A_4B_3$ subunit which is encoded by the Gy_4 gene (Staswick and Nielsen 1983; Kitamura et al. 1984, 1980). The purpose of the work described in this communication was to identify the genetic lesion responsible for the absence of the $A_5A_4B_3$ subunit in seeds in the cultivar Raiden.

* Cooperative research between USDA/ARS and the Indiana Agricultural Experiment Station. This is Journal Paper No. 10,884 from the Purdue Agricultural Experiment Station. Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by USDA/ARS on Purdue University

Abbreviations: bp, base pairs; kb, kilobase pairs; DNase I, deoxyribonuclease I; EDTA, ethylenediaminetetraacetic acid; RNAase TI, ribonuclease TI; SDS, sodium dodecyl sulfate; TLE, 10 mM Tris, 0.1 mM EDTA

Materials and methods

Materials. Restriction enzymes were purchased from either Bethesda Research Laboratories (BRL) or Boehringer Mannheim. T4 ligase was from New England Biolabs. T4 polynucleotide kinase and nick translation kits were from BRL. The α-[^{32}P]dNTPs were from Amersham and γ-[^{32}P]ATP was from ICN Nutritional Biochemicals Inc. Nitrocellulose was purchased from Schleicher and Schuell. Ribonuclease A (RNAase A), ribonuclease T1 (RNAase T1) sucrose and Ficoll were from Sigma Chemical Company. NZY broth was purchased from Gibco Laboratories. Lambda DNA packaging extracts, DNase I and RNasin were purchased from Promega Biotec. Charon 35 bacteriophage were kindly provided by Dr. Jim Mullins, Harvard School of Public Health, Boston, Mass.

Isolation of leaf DNA. Plants of the soybean cultivar Raiden were grown in the greenhouse and leaves harvested when one or two trifoliate leaves had fully emerged. The leaves were immediately frozen in liquid nitrogen and then stored at −80° C until used. DNA was isolated from the leaves by the mini-preparation method of Dellaporta et al. (1983) with two modifications: (1) the leaves were not lyophilized prior to grinding, and (2) following precipitation with cethyl trimethylammonium bromide, the leaf DNA was treated with 30 µg/ml of RNAase A for 30 min at 37° C. The DNA was extracted once with phenol/chloroform and once with chloroform before precipitating with ethanol.

Construction and screening of a genomic library. Raiden leaf DNA was partially digested with *Mbo*I and then size-fractionated on a 5%–20% NaCl gradient. Fractions containing DNA fragments between 14 and 21 kb in size were pooled and the DNA was precipitated with ethanol. Charon 35 arms (Loenon and Blattner 1983) were prepared by a modification of a method described by Maniatis et al. (1982). The *cos* sites of Charon 35 DNA were annealed for 1 h at 42° C in 100 mM Tris-HCl, pH 8.0 and 10 mM $MgCl_2$ (80–120 µg of DNA in 400 µl volume). The annealing mixture was adjusted to 1× ligation buffer and 60–120 Weiss units of T4 ligase were added. The reaction was incubated at 37° C for 1 h and then 15° C for 12 h. The DNA was gently extracted with phenol/chloroform and precipitated with isopropanol. The precipitate was resuspended in TLE and digested to completion with *Bam*HI. The digest was adjusted to 20 mM EDTA and loaded onto a 5%–20% potassium acetate gradient that contained 0.01% Sarkosyl. The sample was spun at 27,000 rpm for 15 h at 20° C in an SW28 Beckman rotor. Following centrifugation the supernatant was aspirated and discarded. The pellet, consisting of annealed arms (30.2 kb), was dried and resuspended in 500 µl TLE. The DNA was extracted with phenol/chloroform, precipitated with isopropanol and resuspended in 50 µl TLE. The purified, annealed arms were aliquoted and stored at −20° C until used.

Charon 35 arms (2.7 µg) and leaf DNA (0.8 µg) were ligated in a 20 µl reaction volume using 9 Weiss units of T4 DNA ligase. The ligation reaction was incubated at 12° C for 15 h and then 68° C for 5 min. The recombinant DNA molecules were either precipitated with ethanol before packaging or were packaged immediately into phage particles. Approximately one-half of a ligation reaction was used for each packaging reaction. The library contained approximately 5×10^5 independent plaque-forming units.

The Raiden genomic library was screened for *Gy*$_4$ clones by the plaque hybridization procedure of Benton and Davis (1977) using as probe the nick-translated insert of a *Gy*$_4$ cDNA clone (pG258). Nitrocellulose filters were prehybridized and hybridized at 42° C in 50% formamide, 5 × SSC, 5 × Denhardt's (0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone), 1% SDS, 50 mM sodium phosphate, pH 6.5, 250 µg/ml denatured calf thymus DNA, 1 mM EDTA, 1 µg/ml poly(dG)-poly(dC), and 0.1 µg/ml linearized denatured pUC8 DNA. After hybridization the filters were washed in 2 × SSC, 0.1% SDS at room temperature and then in 0.1 × SSC, 0.1% SDS at 66° C for 5 h. The filters were autoradiographed with an intersifier screen for 1 to 2 days.

DNA sequencing. Sequencing of DNA was performed as described by Maxam and Gilbert (1980). DNA fragments were end-labeled using either T4 polynucleotide kinase and γ-[^{32}P]ATP (7,000 Ci/mmol) or the Klenow fragment and α-[^{32}P]dNTPs (3,000 Ci/mmol).

RNA isolation. Poly(A)$^+$ RNA was isolated from polysomes of mid-maturation stage embryos of Raiden and CX635-1-1-1 as described by Tumer et al. (1981). Total cellular RNA from mid-maturation stage embryos was isolated by the method of Hall et al. (1978).

Construction of pSP64/248HB and transcription of the antisense Gy$_4$ *probe.* The *Bcl*I−*Hin*dIII fragment of the *Gy*$_4$ cDNA clone pG248 (Scallon et al. 1985) was isolated from an agarose gel and inserted between the *Bam*HI and *Hin*dIII sites of pSP64 to form plasmid pSP64/248HB. The plasmid was linearized by digestion with *Pvu*II and used to synthesize [^{32}P]UTP-labeled glycinin antisense RNA by the procedure of Melton et al. (1984). Antisense transcripts from a 20 µl reaction were precipitated with ethanol and resuspended in 80% formamide, 40 mM Pipes, pH 6.7, 0.4 M NaCl and 1 mM EDTA (hybridization buffer).

RNAase protection. Total cellular RNA (100 µg) or polysome-derived poly(A)$^+$ RNA (1 µg) with carrier tRNA (99 µg), or carrier tRNA alone (100 µg) was dissolved in 30 µl of hybridization buffer. One microliter of ^{32}P-antisense *Gy*$_4$ probe was added to each sample. The mixture was incubated at 85° C for 5 min to denature the RNAs and then incubated at 45° C overnight. Then the samples were sequentially subjected to RNAase T1 treatment (1 h at 30° C), proteinase K digestion, phenol extraction and ethanol precipitation as described by Melton et al. (1984). The precipitate was dissolved in 50 µl of formamide loading buffer, placed in a boiling water bath for 3 min, and 1 µl samples were fractionated in a 6% polyacrylamide sequencing gel. Bands were visualized by autoradiography carried out overnight with intensifying screens. Size standards were used transcripts from *Hae*III-digested pSP65, as well as SP6 control templates from New England Biolabs.

Results and discussion

The cultivar Raiden has been shown by electrophoretic and chromatographic analysis of its seed protein to lack the A_4, A_5 and B_3 polypeptides that are present in most other cultivars (Kitamura et al. 1984; Staswick and Nielsen 1983; Kitamura et al. 1980). Genetic studies have revealed that the absence of all three polypeptides results from a single, recessive allele designated *gy*$_4$ (Kitamura et al. 1984; Scallon et al. 1985).

Southern hybridizations of leaf DNA have shown that the *Gy*$_4$ gene in the wild-type cultivar CX635-1-1-1 is on a 13 kb *Eco*RI fragment and that the gene is present approximately once per haploid genome (Scallon et al. 1985). *Gy*$_5$, a highly homologous glycinin gene which is also present about once per haploid genome, exists on a 9 kb *Eco*RI fragment in CX635-1-1-1. When a mixture of *Gy*$_4$ and *Gy*$_5$ sequences was used as a probe in Southern hybridizations with Raiden genomic DNA, the same banding pattern was observed (e.g. 13 and 9 kb *Eco*RI fragments, data not shown). This indicated that the coding sequence for *gy*$_4$ was present in the Raiden genome on the same size *Eco*RI restriction fragment as in cultivars which produced the subunit. Hence, the failure of the null allele to produce the subunit was not due to a complete deletion of the gene.

Isolation of the Raiden Gy$_4$ *gene*

It was considered likely that the molecular lesion responsible for the absence of the $A_5A_4B_3$ subunit had occurred in the structural gene. To investigate this possibility, a genomic library of Raiden leaf DNA was constructed in the bacteriophage vector Charon 35 (Loenen and Blattner 1983). Approximately 2×10^5 independent plaque-forming units from the library were screened prior to amplification by plaque hybridization. The nick-translated insert in pG258, a cDNA clone for *Gy*$_4$ (Scallon et al. 1985), was used as probe. Wash conditions after hybridization were at stringencies high enough to ensure little or no cross-hybridization to *Gy*$_5$ sequences. Several genomic clones were purified and one of these, λR6, was amplified for further study.

A restriction map of the 16 kb insert of λR6, as well as the position and orientation of the glycinin gene within it, are shown in Fig. 1. The *Hin*dIII and *Xho*I restriction sites located within the gene are common to both *Gy*$_4$ and *Gy*$_5$, but the *Kpn*I site is not present in *Gy*$_5$ (Nielsen et al. in preparation. The presence of the *Kpn*I site provided initial evidence that the insert encoded the *gy*$_4$ rather than the *Gy*$_5$ gene. The 7.7 kb *Bam*HI fragment in λR6, which contained all of the *gy*$_4$ gene, was ligated into the *Bam*HI site of pUC7. This subclone was used for all subsequent investigations into the structure of the Raiden *gy*$_4$ gene.

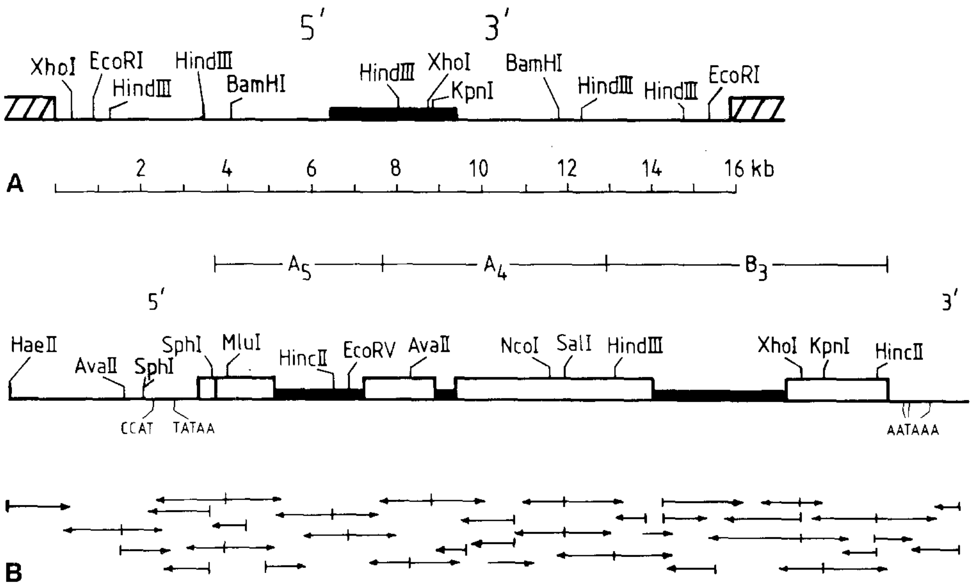

Fig. 1. A Restriction map of the insert of genomic clone λR6. The gy_4 allele is represented by the *black box* in the center of the map. *Hatched boxes* represent the ends of the Charon 35 arms. **B** Structure and sequencing strategy of the gy_4 allele. *White boxes* represent exons and *black boxes* represent introns. The portions of the gene that contain coding information for each of the polypeptides associated with the mature subunit are indicated. *Arrows* represent the strategy used to determine the DNA sequence

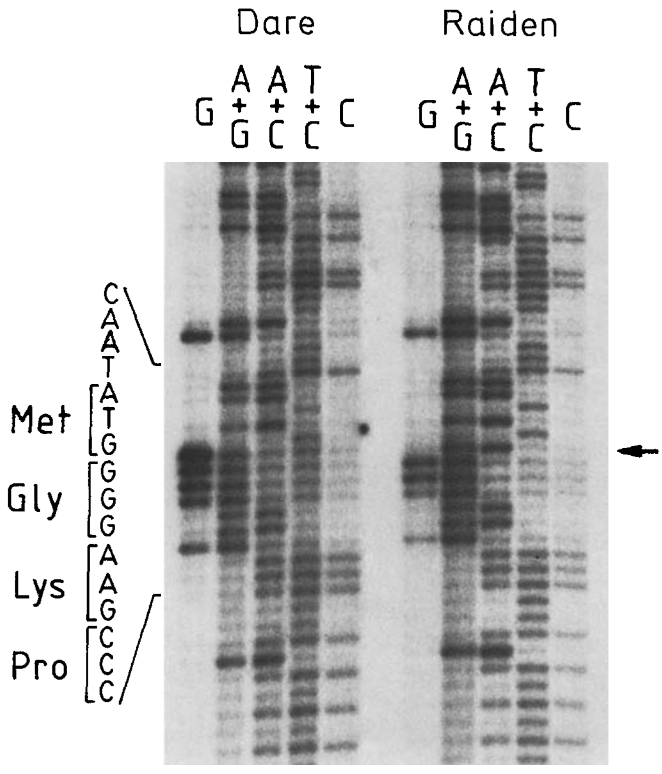

Fig. 2. Autoradiograph of a sequencing gel that reveals the translation initiation codon for the Gy_4 allele from Dare and the corresponding region from the gy_4 null allele in Raiden. The *arrow* marks the single base difference

To test the possibility that the mutation was an insertion or deletion not detected in the Southern hybridization experiments, a more detailed restriction map was made of the gy_4 allele (Fig. 1) and compared with the map of a functional Gy_4 allele from the cultivar Dare. No differences between this map and the one for the functional allele in Dare were observed.

Sequence analysis of the Raiden Gy_4 *gene*

The DNA sequence at or near the 5′ end of the null allele was determined and compared with the corresponding sequence of the Gy_4 gene from Dare. Only one difference was observed in the 200 bp of sequence obtained upstream from the unique *Mlu*I restriction site in the gene. The gene from Raiden lacked the normal translation start codon where a single base change had converted the ATG initiation codon to ATA (Fig. 2).

To determine whether mutations other than the one at the translation start codon existed, the DNA sequence of the entire gene and 600 bp of 5′ flanking DNA was also determined and compared with the functional gene from Dare (Fig. 3). Only two other differences were observed in the 3,500 nucleotides that were sequenced. One was a G/A transition 268 bp by upstream from the transcription initiation site. This mutation was situated approximately 50 bp upstream from the highly conserved region around tandem CCAT sequences common to all of the glycinin genes (Nielsen et al. in preparation). The other was a T/C transition at position 703 in the coding region. The latter base change would result in a leucine for proline substitution. Neither of these two point mutations was likely to result in absence of protein, and it was concluded that the mutation at the initiation codon was probably responsible for the phenotype observed in Raiden, and suggested that nonfunctional gy_4 mRNA might be detected in embryos from Raiden.

Detection of Gy_4 *transcripts*

In order to determine if gy_4 transcripts were present in Raiden embryos, it was necessary to develop a means to distinguish them from transcripts of the highly homologous Gy_5 gene. As earlier characterization of glycinin subunits had revealed a hyper-variable region (Argos et al. 1985), the possibility existed that this region in the two genes might be sufficiently different that an assay to distinguish the transcripts could be developed. A comparison of the coding sequence in the hypervariable regions (Fig. 4) revealed that the major difference between the two genes was a deletion

1 CTATACAATA TAAGATCATA GTACTGACAA AATGCACAGT AAAACAGTTC AAATTGAGAA GGATTCTTAA CACACCATAG TATTTAATAT ATATCTTTAC

101 AGAGACAATT ATGCTGGAGG ATTCAGGCAA AGATTATATA TTGTGGATTT GTTTTTTAAT AATTAACGCA TCATATGAAA GATCGATGAT ATATACTAAT

201 GGTTATAAGA AAAATATTTA ACAGTTTCTA TAACCTTTTT CTTTTATCTT TTACTGTAAT ATTATTTATT TTATTTCACA TTTTTAATCA GCTTATCTCA

301 TTTATAAACG AAATTGTATA AAAATATACA TGATGAACTG AATAGAACAA TATTGG$\frac{T}{A}$CTG ATATTCTCAT ATTGTATAAG AGGATAGACT TTGAGACGCG

401 GAGAATCTGT AGGAGGGGAC CATTCAGAGT GCCTCCAATT TTGGTGTTGT TCATTGTACC ATTGCAAATA TAAACGAAGC ATGCATGCTT ATGTATGAGG

501 TGTAACAAAA TTGGAAACAA TAGCCATGCA AGGTGAAGAA TGTCACAAAC TCAGCAACCC TTATTCATTG ACGTGTCCCT CAGTCACTCT CCTCTCATAC

GlyLysPr oPheThrLeu
601 CTATAAATCA CCACTCCTCA TGTTCTTTCC AATTCACCAA CTCCTTCAAA CTTAATTATT AACACTTCCT TAGTTCAATA T$\frac{A}{G}$GGGAAGCC CTTCACTCTC

SerLeuSerS erLeuCysLe uLeuLeuLeu SerSerAlaC ysPheAlaIl eSerSerSer LysLeuAsnG luCysGlnLe uAsnAsnLeu AsnAlaLeuG
701 TCTCTTTCTT CCCTTTGCTT GCTACTCTTG TCGAGTGCAT GCTTTGCTAT TAGCTCCAGC AAGCTCAACG AGTGCCAACT CAACAACCTC AACGCGTTGG

luProAspHi sArgValGlu SerGluGlyG lyLeuIleGl nThrTrpAsn SerGlnHisP roGluLeuLy sCysAlaGly ValThrValS erLysLeuTh
801 AACCCGACCA CCGCGTTGAG TCCGAAGGTG GTTTGATTCA AACATGGAAC TCTCAACACC CTGAGCTGAA ATGCGCCGGT GTCACTGTTT CCAAACTCAC

rLeuAsnArg AsnGlyLeuH isLeuProSe rTyrSerPro TyrProArgM etIleIleIl eAlaGlnG
901 CCTCAACCGC AATGGCCTCC ACTTGCCATC TTACTCACCT TATCCCCGGA TGATCATCAT CGCCCAAGGT AATCATATAT AAGGAGTGCT TCTAACACAC

1001 ATATCAGAAA GAGTATCACC AGCATTTCTC AGTGTATATT AATCCATTTG TCACCACTTG TTCAAATTTC AACATCACAT TACCATAGAT CATTTACTAA

1101 AGATAATAAT GATTTAAGTA AATAGTATCT CTATAGTAAA TTTTACATGA TTATTTAACT ACAAATTATT ATTATTATAT ATAGAATGAC TTTGTTGACA

1201 TATCAATCAC CTTAAAAGTT TTATTAAGTT ATATATATCA ACTAAGATAT CTGATTAAAT AAAAATGTGA TTGTTTTGTT TGGTGATGAT TGATGTACAG

lyLysGlyAl aLeuGlyVal AlaIleLeuG lyCysProGl uThrPheGlu GluProGlnG luGlnSerAs nArgArgGly SerArgSerG lnLysGlnGl
1301 GGAAAGGAGC ACTTGGAGTT GCAATTC$\frac{T}{C}$AG GATGTCCTGA GACGTTTGAG GAGCCACAAG AACAATCAAA CAGAAGAGGC TCAAGGTCGC AGAAGCAGCA

nLeuGlnAsp SerHisGlnL ysIleArgHi sPheAsnGlu GlyAspValL euValIlePr oProGlyVal ProTyrTrpT hrTyrAsnTh rGlyAspGlu
1401 GCTACAGGAC AGTCACCAGA AGATTCGTCA CTTCAATGAA GGAGACGTAC TCGTGATTCC TCCTGGTGTT CCTTACTGGA CCTATAACAC TGGCGATGAA

ProValValA laIleSerLe uLeuAspThr SerAsnPheA snAsnGlnLe uAspGlnThr ProArg
1501 CCAGTTGTTG CCATCAGTCT TCTTGACACC TCTAACTTCA ATAACCAGCT TGATCAAACC CCTAGGGTAA TTATCAATTC AATTTCATTT ACTATTAACA

ValPheTyr LeuAlaGlyA snProAspIl eGluTyrPro GluThrMetG lnGlnGlnGl
1601 AAAACCATGT TCTCCTCACT TGTTAATTTT TTCACTTTCA GGTATTTTAC CTTGCTGGGA ACCCAGATAT AGAGTACCCA GAGACCATGC AACAACAACA

nGlnGlnLys SerHisGlyG lyArgLysGl nGlyGlnHis GlnGlnGluG luGluGluGl uGlyGlySer ValLeuSerG lyPheSerLy sHisPheLeu
1701 ACAGCAGAAA AGTCATGGTG GACGCAAGCA GGGGCAACAC CAGCAGGAGG AAGAGGAAGA AGGTGGCAGC GTGCTCAGTG GCTTCAGCAA ACACTTCTTG

AlaGlnSerP heAsnThrAs nGluAspIle AlaGluLysL euGlnSerPr oAspAspGlu ArgLysGlnI leValThrVa lGluGlyGly LeuSerValI
1801 GCACAATCCT TCAACACCAA CGAGGACATA GCTGAGAAAC TTCAGTCTCC AGACGACGAA AGGAAGCAGA TCGTGACAGT GGAAGGAGGT CTCAGCGTTA

leSerProLy sTrpGlnGlu GlnGlnAspG luAspGluAs pGluAspGlu AspAspGluA spGluGlnIl eProSerHis ProProArgA rgProSerHi
1901 TCAGCCCCAA GTGGCAAGAA CAACAAGATG AAGATGAAGA TGAAGACGAA GATGATGAAG ATGAACAAAT TCCCTCTCAC CCTCCTCGCC GACCAAGCCA

sGlyLysArg GluGlnAspG luAspGluAs pGluAspGlu AspLysProA rgProSerAr gProSerGln GlyLysArgG luGlnAspGl nAspGlnAsp
2001 TGGAAAGCGT GAACAAGACG AGGACGAGGA CGAAGATGAA GATAAACCTC GTCCTAGTCG ACCAAGCCAA GGAAAGCGTG AACAAGACCA GGACCAGGAC

GluAspGluA spGluAspGl uAspGlnPro ArgLysSerA rgGluTrpAr gSerLysLys ThrGlnProA rgArgProAr gGlnGluGlu ProArgGluA
2101 GAGGACGAAG ATGAAGATGA AGATCAACCT CGCAAGAGCC GCGAATGGAG ATCGAAAAAG ACACAACCCA GAAGACCTAG ACAAGAAGAA CCACGTGAAA

rgGlyCysGl uThrArgAs nGlyValGluG luAsnIleCy sThrLeuLys LeuHisGluA snIleAlaAr gProSerArg AlaAspPheT yrAsnProLy
2201 GAGGATGCGA GACAAGAAAC GGGGTTGAGG AAAATATCTG CACCTTGAAG CTTCACGAGA ACATTGCTCG CCCTTCACGC GCTGACTTCT ACAACCCTAA

```
      sAlaGlyArg IleSerThrL euAsnSerLe uThrLeuPro AlaLeuArgG lnPheGlnLe uSerAlaGln TyrValValL euTyrLys
2301  AGCTGGTCGC ATTAGTACCC TCAACAGCCT CACCCTCCCA GCCCTCCGCC AATTCCAACT CAGTGCCCAA TATGTTGTCC TCTACAAGGT ATGTAATTCA

2401  CCTCATTCAT ATTACTAAGT AATCAACATG AAACTAATAT ACGTACATAC TTACACATCT ACCAGTAATT TTTCCGTGGA TATTCAATTG TCAATTAGTC

2501  TATCTTGAGA AAATTAAGAA ATAAAAAGAA AGCACAAAAG GGAAAAATCT TTATGTCATA AATCATATGA TATAATAATT TAGAAGACAT ATAAAAATGT

2601  CAGTAAGTAT GTTGTAGGGT TGGATTCCTT TAAATGTCAT TAAAATATCA TTTGATATGG GTAATTCTTT AGTGATTCTC TAGGGGTAGT TGAACTGTAA

2701  TGTATTATAA TTGTGCATTG ATTTTTATGA GTTACTTTAA CATGTCAATG AAGACTTATT TGATAATAAT TATAGTTACT TGTTGGTTCT ACTACTTTTA

                                                                                                       A snGlyIleTy
2801  ATAAAAAAAT AATAAAAATA TTGGTGTAAA TATATAATAT ATAATAATAA TGATGATGAT ACGTAACACA TGTTATTATA TCCATGCAGA ATGGAATTTA

      rSerProHis TrpAsnLeuA snAlaAsnSe rValIleTyr ValThrArgG lyGlnGlyLy sValArgVal ValAsnCysG lnGlyAsnAl aValPheAsp
2901  CTCTCCACAT TGGAATCTGA ATGCAAACAG TGTGATCTAT GTGACTCGAG GACAAGGAAA GGTTAGAGTT GTGAACTGCC AAGGGAATGC AGTGTTCGAC

      GlyGluLeuA rgArgGlyGl nLeuLeuVal ValProGlnA snPheValVa lAlaGluGln AlaGlyGluG lnGlyPheGl uTyrIleVal PheLysThrH
3001  GGTGAGCTTA GGAGGGGACA ATTGCTGGTG GTACCACAGA ACTTCGTGGT GGCGGAGCAA GCCGGAGAAC AAGGATTCGA ATACATAGTA TTCAAGACAC

      isHisAsnAl aValThrSer TyrLeuLysA spValPheAr gAlaIlePro SerGluValL euAlaHisSe rTyrAsnLeu ArgGlnSerG lnValSerGl
3101  ACCACAACGC AGTCACTAGC TACTTGAAGG ATGTGTTTAG GGCAATTCCC TCAGAGGTTC TTGCCCATTC TTACAACCTT CGACAGAGTC AAGTGTCTGA

      uLeuLysTyr GluGlyAsnT rpGlyProLe uValAsnPro GluSerGlnG lnGlySerPr oArgValLys ValAla---
3201  GCTTAAGTAT GAAGGAAATT GGGGTCCTTT GGTCAACCCT GAGTCTCAAC AAGGCTCACC CCGTGTTAAA GTCGCATAAA TGACAAGCAT GATGGTGTGA

3301  GGATGAGGCC ATCTTATGAA ATAATAACAA ATAAATAAAT TTTGTATGAT AATAAAAAGT ATGGCCCATG TACCATCCCA GCGAGCCTAT GTTTATATCT

3401  GAGTGGCGTT GTACCTTTCA ATCGCCTTAA TAAAATGTCA GTCTTCACGT TTTGTCTTTA TTCTGTGTTT ATTTTCTTTT TTGTGGGCAA GCTAGCTTTT

3501  ATCTACTTTT AAATGAGTAA TACTGATTAT ATGTTTACTG GGG
```

Fig. 3. DNA sequences and deduced amino acid sequences of the Gy_4 genes from Dare and Raiden. Nucleotides in Dare are written below the Raiden sequence only where they differ from Raiden (*underlined*). The positions of the three polypeptide components of this glycinin subunit are indicated

```
Gy4   541 GATCAAACCCCTAGGGTATTTACCTTGCTGGGAACCCAGATATAGAGTACCCAGAGACCATGCAACAACAACAACAGCAGAAAAGTCATGGTGGACGCA
Gy5   541         A   C  A                                C  T  C                 G  G        G

Gy4   641 AGCAGGGGCAACACCAGCAGGAGGAAGAGGAAGAAGGTGGCAGCGTGCTCAGTGGCTTCAGCAAACACTTCTTGGCACAATCCTTCAACACCAACGAGGA
Gy5   641                GA   C  ---                 T                       T     A

Gy4   741 CATAGCTGAGAAACTTCAGTCTCCAGACGACGAAAGGAAGCAGATCGTGACAGTGGAAGGAGGTCTCAGCGTTATCAGCCCCAAGTGGCAAGAACAACAA
Gy5   738   C              G         T                             G     C                                 ---

Gy4   841 GATGAAGATGAAGATGAAGACGAAGATGATGAAGATGAACAAATTCCCTCTCACCCTCCTCGCCGACCAAGCCATGGAAAGCGTGAACAAGACGAGGACG
Gy5   835 ---     C      C           C A   T   G  GG CG      T T     A  A                   A    G T

Gy4   941 AGGACGAAGATGAAGATAAACCTCGTCCTAGTCGACCAAGCCAAGGAAAGCGTGAACAAGACCAGGACCAGGACGAGGACGAAGATGAAGATGAAGATCA
Gy5   932            A      C          GA  AC  TCCA  G--------------------------------------------------------

Gy4  1032 ACCTCGCAAGAGCCGCGAATGGAGATCGAAAAAGACACAACCCAGAAGACCTAGACAAGAAGAACCACGTGAAAGAGGATGCGAGACAAGAAACGGGGTT
Gy5   974 -------------------------------------     A  C  G  CGA    C            G          TC   T     T

Gy4  1132 GAGGAAAATATCTGCACCTTGA
Gy5  1040           T      A
```

Fig. 4. Comparison of the *BclI – HindIII* fragments from Gy_4 and Gy_5. *Dashes* represent gaps introduced into the sequence to permit maximal alignment. The Gy_4 sequence is from cDNA clone pG248 produced from the cultivar CX635-1-1-1. The Gy_5 sequence is from a cDNA clone from the cultivar Bonminori reported by Fukazawa et al. (1985)

of 93 bp in Gy_5 compared with the analogous part in gy_4. Hence, production of an antisense RNA probe complementary to this region of gy_4 would provide a means to distinguish between the gy_4 and Gy_5 transcripts. Cross-hybridization of antisense gy_4 RNA to Gy_5 transcripts would result in a single-stranded loop at the position of the deletion. RNAase T1 digestion of single-stranded RNA in the loop would generate protected regions of probe that would be shorter than those between probe and gy_4 RNA. The difference in size of protected fragments would be sufficiently different due to the 93 bp deletion in Gy_5 that they could easily be distinguished electrophoretically.

Plasmid pSP64/248HP was constructed for the production of ^{32}P-labeled RNA probe. It contained the hypervariable region in Gy_4 that was encoded on a 630 bp *Bcl*I–*Hin*dIII fragment from cDNA clone pG248 (Scallon et al. 1985). The template was prepared by linearizing the plasmid construct by digestion at a unique *Pvu*II site located 190 bp downstream from the Gy_4 insert. The nascent probe produced by SP6 RNA polymerase from the template was therefore 820 bases long. The extra 190 bp on the probe provided a means to distinguish between probe molecules not digested with RNAase T1 and shorter ones which had been digested but were protected by either gy_4 or Gy_5 mRNAs.

Total cellular RNA from Raiden protected 630 bases of the probe from RNAase T1 digestion (Fig. 5). This result clearly indicated that the gy_4 allele in Raiden was transcribed in vivo. As anticipated, total cellular RNA from the control cultivar with a functional allele also protected 630 bases of the probe. Total cellular RNA from both cultivars protected fragments of about 430 bases which were due to protection by Gy_5 RNAs.

Different results were obtained when the antisense probe was incubated with polysomal poly(A)$^+$ RNA. A strong signal was obtained when poly(A)$^+$ RNA isolated from a cultivar with a functional Gy_4 allele was used to protect the probe, whereas a very weak signal was observed in the case of poly(A)$^+$ RNA from the cultivar with the null allele. The doublet at the 630 base position probably was the result of partial protection by the vector near the ends of the 630 base conserved region (Melton et al. 1984). This result was consistent with a Northern hybridization experiment in which a Gy_4-specific probe produced a strong signal when hybridized to CX635-1-1-1 poly(A)$^+$ RNA, but a near undetectable signal when hybridized to the same amount of Raiden poly(A)$^+$ RNA (Scallon 1986). These data indicated that gy_4 mRNAs from the null allele in Raiden are not effectively retained in polysomes like the functional mRNAs from CX635-1-1-1.

Eukaryotic ribosomes almost always initiate translation at the first available AUG from the 5′ end of the mRNA (Baim et al. 1985). The first available AUG that occurred in the Raiden gy_4 transcript would be 60 bp downstream from the mutant initiation site. This AUG would be in an "incorrect" reading frame and would be followed four codons later by a translation stop codon. Other potential AUG codons located even further downstream were also out of frame and followed by stop codons. It is likely that the lack of a strong gy_4 signal in poly(A)$^+$ RNA from the null allele reflects termination of translation near the beginning of the transcript und the subsequent immediate release of the defective message from the ribosome.

The signals at 430 bases and smaller (Fig. 5) are due to protection of the probe by Gy_5 mRNAs. The heterogenity observed in the signal at about 430 bases is considered a result of the poor homology that exists between Gy_4 and probe at both ends of the unpaired region (Fig. 4). A difference can be seen between the two cultivars in the intensity

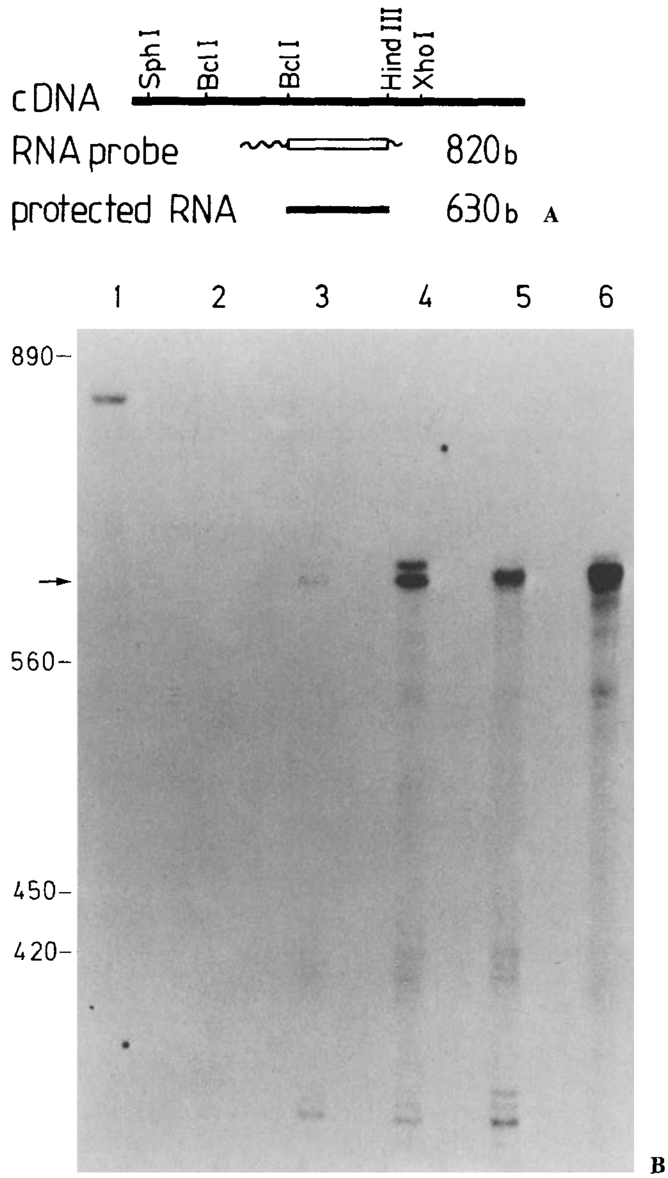

Fig. 5A, B. Ribonuclease protection analysis of RNA from Raiden and CX635-1-1-1 seeds. **A** Schematic representation of the Gy_4-specific RNA probe generated by SP6 polymerase. The 830 base probe contains sequence transcribed from a segment of Gy_4 cDNA (*open box*), as well as sequence derived from the SP6 vector (*wavy lines*). A fragment of approximately 630 bases is expected from protection by gy_4 mRNA, while smaller ones (430 bases or less) were expected from protection by Gy_5 mRNA. **B** RNAase protection analysis. A Gy_4-specific ^{32}P-labeled single-stranded RNA probe was annealed to one of four different RNA fractions: lane 2, tRNA from yeast; lane 3, poly(A)$^+$ RNA isolated from Raiden polysomes; lane 4, poly(A)$^+$ RNA isolated from CX635-1-1-1 polysomes; lane 5, total RNA from Raiden; and lane 6, total RNA from CX635-1-1-1. Samples were treated with ribonuclease T1 and the products resolved in a 6% polyacrylamide sequencing gel. Lane 1 contains untreated probe. The *arrow* indicates the position of bands that result from protection by gy_4 mRNA. Numbers to the left of the panel indicate position of size standards in bases of these signals in the total RNA lanes relative to the intensity of the 630 base signals. Since the gy_4 signal from Raiden total RNA was weaker than the one from CX635-1-1-1 total RNA, it was possible that transcripts from gy_4 comprised a smaller proportion of the total RNA population in the cultivar with the null allele compared with the control. This decrease may have reflected varietal differences, although it more likely resulted from a higher turnover rate of gy_4 mRNA from the null allele. The latter would be expected since a high proportion of mRNA from fully functional alleles would be protected from ribonuclease digestion in the cytoplasm by virtue of its association with polysomes.

The data described in this report provide a simple and logical explanation for the absence of the $A_5A_4B_3$ subunit in seeds from Raiden. This subunit is encoded by a gene that has sustained a point mutation in its translation initiation codon. As a result, transcripts of this gene are defective and are not translated in the normal reading frame. Because of translation stop codons encountered shortly after the next AUG start codon, only low levels of gy_4 mRNA remain associated with polysomes.

The plant null alleles previously characterized at the molecular level have been found to be due to deletions and insertions that affect transcription. The situation with the gy_4 null allele is in contrast to those null alleles in that a translational defect appears responsible for the null phenotype. It is likely that premature termination of translation caused by point mutations will not be an uncommon feature of null alleles.

Acknowledgements. This work was supported in part by grants from the USDA Competitive Grants Program and the American Soybean Association Research Foundation

References

Argos P, Narayana SVL, Nielsen NC (1985) Structural similarity between legumin and vicilin storage proteins from legumes. EMBO J 4:1111–1117

Baim SB, Goodhue CT, Pietras DF, Eustice DC, Labhard M, Friedman LR, Hampsey DM, Stiles JI, Sherman F (1985) Rules of translation: studies with altered forms of the yeast CYC1 gene. In: Calender C, Gold L (eds) Sequence specificity in transcription and translation. Alan R Liss, New York, pp 351–362

Bennetzen JL, Swanson J, Taylor WC, Freeling M (1984) DNA insertion in the first intron of maize Adh1 affects message levels: cloning of progenitor and mutant Adh1 alleles. Proc Natl Acad Sci USA 81:4125–4128

Benton WD, Davis RW (1977) Screening lgt recombinant clones by hybridization to single plaques in situ. Science 196:180–182

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA mini-preparation: version II. Plant Mol Biol Rep 1:19–21

Fukazawa C, Momma T, Hirano H, Harada K, Udaka K (1984) Glycinin A_3B_4 mRNA: cloning and sequencing of double-stranded cDNA complementary to a soybean storage protein. J Biol Chem 260:6234–6245

Goldberg RB, Hoschek G, Vodkin LO (1983) An insertion sequence blocks the expression of a soybean lectin gene. Cell 33:465–475

Hall TC Ma Y, Buchbinder BU, Pyne JW, Dun SM, Bliss FA (1978) Messenger RNA for G1 protein of French bean seeds: cell free translation and product characterization. Proc Natl Acad Sci USA 75:3196–3200

Kitamura K, Toyokawa Y, Harada K (1980) Polymorphism of glycinin in soybean seeds. Phytochem 19:1841–1843

Kitamura K, Davies CS, Nielsen NC (1984) Inheritance of alleles for Cgy1 and Gy4 storage protein genes in soybean. Theor Appl Genet 68:253–257

Ladin BF, Doyle JJ, Beachy RN (1984) Molecular characterization of a deletion mutation affecting the A′-subunit of B-conglycinin of soybean. J Mol Appl Genet 2:372–380

Loenon WAM, Blattner FR (1983) Lambda Charon vectors (Ch 32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. Gene 26:171–179

Maniatis F, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, New York

Maxam AM, Gilbert W (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560

Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res 12:7035–7056

Nielsen NC (1984) The chemistry of legume storage proteins. Philos Trans R Soc Lond B 304:9921–9926

Scallon BJ (1986) Identification of group-II glycinin genes and the molecular characterization of a group-II glycinin null allele. PhD Thesis, Purdue University, West Lafayette, IN 47907

Scallon BJ, Thanh VH, Floener LA, Nielsen NC (1985) Identification and characterization of DNA clones encoding group-II glycinin subunits. Theor Appl Genet 70:510–519

Staswick PE, Nielsen NC (1983) Characterization of a soybean cultivar lacking certain glycinin subunits. Arch Biochem Biophys 223:1–8

Taylor LP, Walbot V (1985) A deletion adjacent to the maize transposable element Mu-1 accompanies loss of Adhl expression. EMBO J 4:869–876

Tumer N, Thanh VH, Nielsen NC (1981) Purification and characterization of mRNA from soybean seeds. J Biol Chem 256:8756–8760

Communicated by R.B. Goldberg

Received November 29, 1986